

# Consensus of rankings

Zhiwei Lin<sup>\*1</sup>, Yi Li<sup>†2</sup>, and Xiaolian Guo<sup>‡1</sup>

<sup>1</sup>School of Computing and Mathematics, Ulster University,  
United Kingdom

<sup>2</sup>Division of Mathematics, SPMS, Nanyang Technological  
University

April 28, 2017

## Abstract

Rankings are widely used in many information systems. In information retrieval, a ranking is a list of ordered documents, in which a document with lower position has higher ranking score than the documents behind it.

This paper studies the consensus measure for a given set of rankings, in order to understand the degree to which the rankings agree and the extent to which the rankings are related. The proposed approach, without the need for pairwise comparison between rankings, allows to measure the consensus in a set of rankings, with respect to the length of common patterns, the number of common patterns for a given length, and the number of all common patterns. The experiments show that the proposed approach can be used to compare the search engines in terms of closeness of the returned results when semantically related key words are sent to them.

## 1 Introduction

Rankings are widely used in information retrieval, recommender and decision making systems [1, 4, 5]. A ranking is a list of ordered items, in which an item

---

<sup>\*</sup>z.lin@ulster.ac.uk

<sup>†</sup>yili@ntu.edu.sg

<sup>‡</sup>guo-x3@email.ulster.ac.uk

with higher ranking score is placed ahead of the items with lower ranking scores. Information retrieval is a typical example of using rankings, where a ranking result is returned as a sequence of ordered documents after a query is issued. The auto suggestions, important part of search engines nowadays, produce a ranking list of search terms after one or two keystrokes, in order to provide better search experience.

The problem of consensus measure of rankings is the study of quantifying the degree to which the rankings agree according to certain common patterns. Formally, given a set  $\mathcal{R} = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$  of  $N$  rankings, the consensus measure is to quantify and evaluate the commonality hidden in the rankings. The consensus measure of rankings is very useful and important in many information systems. For example, for human, we know that the terms of "007", "James Bond" and many others are conceptually close and related, and they are often used interchangeably. After the queries of those terms are sent to a search engine, we would expect that the search engine return similar results to some extent. Therefore, the degree of "conceptual closeness and relatedness" has to be quantified or measured, in order to compare different search engines, with a given set of related search terms.

Some pairwise comparison approaches have been used, including the Kendall  $\tau$  and Spearman  $\rho$  [3, 6]. Both evaluate the dissimilarity between two rankings  $\mathbf{r}_x$  and  $\mathbf{r}_y$ , i.e.,  $\tau(\mathbf{r}_x, \mathbf{r}_y)$  and  $\rho(\mathbf{r}_x, \mathbf{r}_y)$ . However, it has been proved that both of them are not a valid consensus measure as they violate some of the properties of the consensus measure, as pointed out by Elzinga et al. [2]. This paper investigates the consensus measure of rankings for the degree to which the rankings agree, without the pairwise comparison in Kendall  $\tau$  or Spearman  $\rho$ .

The contribution of the paper lies in the algorithms for the consensus measure with (1)  $\kappa(\mathcal{R})$  – the number of common patterns within the rankings of  $\mathcal{R}$ ; (2)  $\kappa_p(\mathcal{R})$  – the number of common patterns of lengths  $p$  in  $\mathcal{R}$ ; (3)  $\ell(\mathcal{R})$  – the length of the longest common patterns in  $\mathcal{R}$ ; and (4)  $\hat{\kappa}_p(\mathcal{R})$  and  $\hat{\kappa}(\mathcal{R})$  – an approach to measuring the degree of consensus, based on not only the number of sub-patterns and the lengths of the common patterns, but also the gaps of the items in the original rankings. We did experiments to show how the proposed can be used for comparing different search engines results.

## 2 Ranking sequences and patterns

A ranking is often presented as a sequence of ordered items, in which an item with lower position in the sequence is more preferred than the items behind it [2]. Formally, let  $\Sigma = \{\sigma_1, \dots, \sigma_n\}$  be a set of  $n$  items, a ranking  $\mathbf{r}$  is an

ordered sequence  $\mathbf{r} = (\sigma_{i_1}, \sigma_{i_2}, \dots, \sigma_{i_m})$  of  $m$  ( $m \leq n$ ) distinct items drawn from  $\Sigma$ . The length of  $\mathbf{r}$  is denoted by  $|\mathbf{r}|$ . For notational simplicity, we shall simply write a ranking as a sequence of  $\mathbf{r} = \sigma_{i_1} \sigma_{i_2} \dots \sigma_{i_m}$  in the rest of the paper.

For a ranking  $\mathbf{r} = r_1 \dots r_k$ , where  $r_j \in \Sigma$  for  $1 \leq j \leq k$ , we can define the embedded patterns with respect to *subsequences*. A sequence  $\mathbf{r}' = r'_1 \dots r'_m$  is called a *subsequence* of  $\mathbf{r}$ , denoted by  $\mathbf{r}' \sqsubseteq \mathbf{r}$ , if  $\mathbf{r}'$  can be obtained by deleting  $k - m$  items from  $\mathbf{r}$ .  $\mathbf{r}' \not\sqsubseteq \mathbf{r}$  is used to denote that  $\mathbf{r}'$  is not a subsequence of  $\mathbf{r}$ . For example,  $bde \sqsubseteq abcde$ , and  $bac \not\sqsubseteq abcde$ .

A ranking sequence with no items is an *empty sequence*. We use  $\mathcal{S}(\mathbf{r})$  to denote the set of all possible non-empty subsequences of  $\mathbf{r}$ , and use  $\mathcal{S}(\mathbf{r} : \sigma)$  to denote the set of all subsequences terminating with  $\sigma$  (i.e., the final item of the subsequence is  $\sigma$ ). Then,  $\mathcal{S}(\mathbf{r}) = \bigcup_{\sigma} \mathcal{S}(\mathbf{r} : \sigma)$ , where the union is taken over the set of the sequence items of  $\mathbf{r}$ . Alternatively,  $\mathcal{S}(\mathbf{r})$  can be partitioned into subsets  $\mathcal{S}_p(\mathbf{r})$ , in which  $\mathcal{S}_p(\mathbf{r})$  consists of all subsequences of length  $p$ . For example, if  $\mathbf{r} = abcde$ , then  $\mathcal{S}(\mathbf{r} : d) = \{d, ad, bd, cd, abd, acd, bcd, abcd\}$ , where each subsequence terminates with  $d$ , and  $\mathcal{S}_3(\mathbf{r}) = \{abc, abd, abe, bcd, bce, bde, cde\}$ , in which each subsequence has length 3.

The degree to which rankings agree lies in the common patterns or subsequences which are embedded by the rankings. Given a set of  $N$  rankings  $\mathcal{R} = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$ , consider  $\mathcal{S}(\mathcal{R}) = \mathcal{S}(\mathbf{r}_1) \cap \dots \cap \mathcal{S}(\mathbf{r}_N)$ , each element  $x \in \mathcal{S}(\mathcal{R})$  is a *common subsequence* of  $\mathbf{r}_1, \dots, \mathbf{r}_m$ , for which we also use the notation  $x \sqsubseteq \mathcal{R}$ . Similar to  $\mathcal{S}(\mathbf{r} : \sigma)$  and  $\mathcal{S}_p(\mathbf{r})$ , we also define  $\mathcal{S}(\mathcal{R} : \sigma)$  to denote the subsets of all common subsequences terminating with  $\sigma$ , and  $\mathcal{S}_p(\mathcal{R})$  to denote the subsets of all common subsequences of length  $p$ , respectively. Therefore, it holds that  $\mathcal{S}(\mathcal{R}) = \bigcup_{\sigma \in \Sigma} \mathcal{S}(\mathcal{R} : \sigma)$  and  $\mathcal{S}(\mathcal{R}) = \bigcup_{1 \leq p \leq l} \mathcal{S}_p(\mathcal{R})$ , where  $l = \min\{|\mathbf{r}| : \mathbf{r} \in \mathcal{R}\}$ .

It is clear that  $\mathcal{S}(\mathcal{R})$  accommodates all common patterns (subsequences), which are subsumed by each ranking  $\mathbf{r} \in \mathcal{R}$ . Let  $\kappa(\mathcal{R})$  denote the number of all common subsequences of  $\mathcal{R}$ , i.e.,  $\kappa(\mathcal{R}) = |\mathcal{S}(\mathcal{R})|$ . The more common patterns  $\mathcal{S}(\mathcal{R})$  has or the bigger  $\kappa(\mathcal{R})$  is, the better consensus  $\mathcal{R}$  has. We can also define  $\kappa_p(\mathcal{R}) = |\mathcal{S}_p(\mathcal{R})|$ , in order to measure the consensus in  $\mathcal{R}$ , with respect to the number of subsequences of a given length  $p$ . The length of the longest common subsequences of rankings in  $\mathcal{R}$  is denoted by  $\ell(\mathcal{R})$  and  $\ell(\mathcal{R}) = \max\{|z| : z \in \mathcal{S}(\mathcal{R})\}$ .

### 3 Consensus measure of rankings

The proposed  $\kappa(\mathcal{R})$ ,  $\kappa_p(\mathcal{R})$  and  $\ell(\mathcal{R})$  enable us to examine the extent of consensus within  $\mathcal{R}$  from three different perspectives. In order to build a

consensus measure for a set of rankings  $\mathcal{R}$ , we start with an example of calculating  $\kappa_p(\mathcal{R})$  for  $\mathcal{R} = \{\mathbf{r}_1 = bdcea, \mathbf{r}_2 = abcde, \mathbf{r}_3 = bdce\}$ . Pick<sup>1</sup>  $\mathbf{r}_2 \in \mathcal{R}$  and form a lower triangle matrix  $\mathbf{A}$  of size  $|\mathbf{r}_2| \times |\mathbf{r}_2|$ , where for  $i \geq j$ ,  $A_{ij} = 1$  if the  $i^{\text{th}}$  item and the  $j^{\text{th}}$  item both occur in the same order in all rankings in  $\mathcal{R}$ , and  $A_{ij} = 0$  otherwise. The matrix  $\mathbf{A}$  we obtain is as follows.

$$\mathbf{A}_{5 \times 5} = \begin{matrix} & \begin{matrix} a & b & c & d & e \end{matrix} \\ \begin{matrix} a \\ b \\ c \\ d \\ e \end{matrix} & \begin{pmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 & 0 \\ 0 & 1 & 1 & 1 & 1 \end{pmatrix} \end{matrix} \quad (1)$$

In the matrix,  $A_{43} = 0$  because  $cd$  is not a subsequence of  $\mathbf{r}_1$  and  $\mathbf{r}_3$ , and all the values in the first column are 0 because  $a$  does not occur in  $\mathbf{r}_3$ .

The matrix  $\mathbf{A}$  induces a directed acyclic graph (DAG)  $G = (V, E)$  on the diagonal elements of  $\mathbf{A}$  as follows, where  $V = \{A_{11}, \dots, A_{55}\}$  is the set of vertices and  $E$  the set of edges. For  $1 \leq i, j \leq |\mathbf{r}|$ , we add an edge from  $A_{ii}$  to  $A_{jj}$  if the following conditions all hold: (1)  $i < j$ ; (2)  $A_{ii} = A_{jj} = 1$ ; (3) if  $A_{ji} \neq 0$ . The induced DAG for our matrix  $\mathbf{A}$  is shown in Figure 1. There are five directed edges, i.e.,  $E = \{(A_{22}, A_{33}), (A_{22}, A_{44}), (A_{22}, A_{55}), (A_{33}, A_{55}), (A_{44}, A_{55})\}$ , suggesting that the number of common subsequences of length 2 is  $\kappa_2(\mathcal{R}) = 5$  and that the common subsequences of length 2 are  $bc, bd, be, ce, de$ , all occurring in  $\mathbf{r}_1, \mathbf{r}_2$  and  $\mathbf{r}_3$ . Note that  $A_{33}$  and  $A_{44}$  have both incoming and outgoing edges. If we traverse from  $A_{22}$ , via either  $A_{33}$  or  $A_{44}$ , to node  $A_{55}$ , we can obtain 3-long subsequences  $bce, bde$  and  $\kappa_3(\mathcal{R}) = 2$ . Therefore, together with the 4 singletons  $b, c, d, e$  ( $\kappa_1(\mathcal{R}) = 4$ ), we have  $\kappa(\mathcal{R}) = \kappa_1(\mathcal{R}) + \kappa_2(\mathcal{R}) + \kappa_3(\mathcal{R}) = 4 + 5 + 2 = 11$ .

This process of finding patterns with various lengths not only allows us to calculate  $\kappa_p(\mathcal{R})$ , but also makes it easy for us to calculate the number of all common patterns  $\kappa(\mathcal{R})$  and the length of the longest common subsequences  $\ell(\mathcal{R})$ . For simpler presentation of our algorithm, we introduce the following indicator function

$$\delta(x) = \begin{cases} 1 & x > 0; \\ 0 & \text{otherwise.} \end{cases} \quad (2)$$

Now we present the lemma for calculating  $\kappa_p(\mathcal{R})$ .

**Lemma 1.** *Given a set  $\mathcal{R} = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$  of  $N$  rankings over a set of items  $\Sigma$ , where each ranking  $\mathbf{r}_k = r_{k_1} \dots r_{k_m}$  is naturally associated with a mapping*

---

<sup>1</sup>The algorithm will work regardless of the choice of ranking.

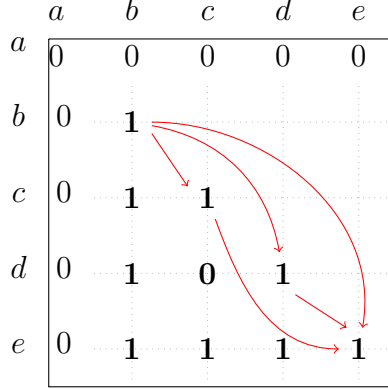


Figure 1: The DAG of the matrix  $\mathbf{A}$  (in Equation (1)) with directed edges  $(A_{ii}, A_{jj})$ , from  $A_{ii}$  to  $A_{jj}$ , where  $i < j$ , if  $A_{ji} \neq 0$ .

function  $\eta_k : \Sigma \rightarrow [0, |\Sigma|]$  defined as

$$\eta_k(\sigma) = \begin{cases} 0, & \sigma \not\sqsubseteq \mathbf{r}_k; \\ j, & \sigma = r_{k_j}. \end{cases} \quad (3)$$

Let  $\mathbf{r}_x = r_{x_1}r_{x_2}\cdots r_{x_n}$  be an arbitrary ranking from  $\mathcal{R}$ ,  $n = |\mathbf{r}_x|$ , and  $\mathbf{A} = (A_{ij})_{n \times n}$ , where

$$A_{ij} = \begin{cases} 0, & i < j; \\ \prod_{k=1}^N \delta(\eta_k(r_{x_i})), & i = j; \\ \prod_{k=1}^N \delta(\eta_k(r_{x_i}) - \eta_k(r_{x_j})) A_{ii}A_{jj}, & i > j. \end{cases} \quad (4)$$

and let  $\mathbf{K} = (K_{pi})_{n \times n}$ , where for  $p > i$ ,  $K_{pi} = 0$ , and for  $p \leq i$

$$K_{pi} = \begin{cases} A_{ii}, & p = 1; \\ \sum_{j=1}^{i-1} K_{p-1j}A_{ij}, & p > 1. \end{cases} \quad (5)$$

Then,  $\kappa_p(\mathcal{R}) = \sum_{1 \leq i \leq n} K_{pi}$ .

*Proof.* Recall that  $\mathcal{S}_p(\mathcal{R})$ , by definition, can be decomposed into disjoint union as

$$\mathcal{S}_p(\mathcal{R}) = \bigcup_{\sigma \in \Sigma} \mathcal{S}_p(\mathcal{R} : \sigma).$$

Note that  $\mathcal{S}_p(\mathcal{R} : \sigma) = \emptyset$  if  $\sigma \not\sqsubseteq \mathbf{r}_x$ , the decomposition above is in fact

$$\mathcal{S}_p(\mathcal{R}) = \bigcup_{1 \leq i \leq n} \mathcal{S}_p(\mathcal{R} : r_{x_i}).$$

Recall that  $\kappa_p(\mathcal{R}) = |\mathcal{S}_p(\mathcal{R})|$ , it suffices to show that  $K_{pi} = |\mathcal{S}_p(\mathcal{R} : r_{x_i})|$ .

Let  $y = r_{x_1}r_{x_2} \cdots r_{x_i}$  be the possible longest subsequence of  $\mathbf{r}_x$  which terminates with  $r_{x_i}$ . Since  $|y| = i$ , we know that any common subsequences terminating with  $r_{x_i}$  has length at most  $i$ . Therefore for  $p > i$  there are no  $p$ -long subsequences terminating with  $r_{x_i}$ , and thus  $\mathcal{S}_p(\mathcal{R} : r_{x_i}) = \emptyset$ , which agrees with our definition that  $K_{pi} = 0$  for  $p > i$ .

For  $p \leq i$  we shall prove by induction on  $p$ . Consider the base case where  $p = 1$ . If  $\mathcal{S}_p(\mathcal{R} : r_{x_i}) \neq \emptyset$ , singleton subsequence  $r_{x_i}$  is the only subsequence in  $\mathcal{S}_p(\mathcal{R} : r_{x_i})$ , that is,  $|\mathcal{S}_p(\mathcal{R} : r_{x_i})| = 1$ . The fact that  $r_{x_i} \in \mathcal{S}_p(\mathcal{R} : r_{x_i})$  also implies that  $r_{x_i} \sqsubseteq \mathcal{R}$ , which means that  $r_{x_i} \sqsubseteq \mathbf{r}_k$  for all  $k$  and thus  $A_{ii} = 1$  by definition. If  $\mathcal{S}_p(\mathcal{R} : r_{x_i}) = \emptyset$ , then there exists  $k$  such that  $r_{x_i} \not\sqsubseteq \mathbf{r}_k$  and thus  $\delta(\eta_k(r_{x_i})) = 0$  and  $E_{ii} = 0$ . Therefore it indeed holds that  $K_{1i} = A_{ii}$ .

Now consider the inductive step for  $p > 1$ . Observe that each  $y \in \mathcal{S}_p(\mathcal{R} : r_{x_i})$  has the form  $y = zr_{x_i}$  for some  $z \in \mathcal{S}_{p-1}(\mathcal{R} : r_{x_j})$ , provided that  $r_{x_j}\sigma_{x_i} \sqsubseteq \mathcal{R}$ . Clearly if  $j \geq i$ , then  $\sigma_{x_j}\sigma_{x_i} \not\sqsubseteq \mathbf{r}_x$  and hence  $\sigma_{x_j}\sigma_{x_i} \not\sqsubseteq \mathcal{R}$ . Therefore, we obtain that

$$\mathcal{S}_p(\mathcal{R} : \sigma_{x_i}) = \bigcup_{1 \leq j < i} \{zr_{x_i} : (z \in \mathcal{S}_{p-1}(\mathcal{R} : r_{x_j})) \wedge (r_{x_j}r_{x_i} \sqsubseteq \mathcal{R})\}. \quad (6)$$

Note that the second constraint  $r_{x_j}r_{x_i} \sqsubseteq \mathcal{R}$  is equivalent to

$$(\eta_k(\sigma_{x_i}) - \eta_k(\sigma_{x_j})) A_{ii}A_{jj} > 0$$

for all  $\mathbf{r}_k \in \mathcal{R}$ , which agrees with the definition of  $A_{ij}$ , that is,  $r_{x_j}r_{x_i} \sqsubseteq \mathcal{R}$  if and only if  $A_{ij} = 1$ . It is the induction hypothesis that  $K_{p-1,i} = |\mathcal{S}_{p-1}(\mathcal{R} : r_{x_i})|$ . It follows from (6) that

$$K_{pi} = \sum_{j=1}^{i-1} |\mathcal{S}_{p-1}(\mathcal{R} : \sigma_{x_j})| A_{ij} = \sum_{j=1}^{i-1} K_{p-1,j} A_{ij},$$

which is exactly the recurrence relation of  $K_{pi}$  in our definition  $\square$

Lemma 1 shows that once we calculate  $\mathbf{K}$ , we simply sum up each row in order to obtain  $\kappa_p(\mathcal{R})$ . Algorithm 1 shows the pseudocode for computing  $\mathbf{K}$ , which uses an additional matrix  $\mathbf{M}$  (on Line 2) to record the position for  $\mathbf{r}_{x_i}$  in  $\mathbf{r}_k$  for  $1 \leq k \leq N$ . The calculation of matrix  $\mathbf{M}$  takes  $O(Nn)$  time. The loop starting from Line 4 runs in time  $O(Nn^2)$  as calculating each individual  $A_{ij}$  loops over all  $N$  rows of  $\mathbf{M}$ . The loop starting from Line 14 to calculate  $\mathbf{K}$  runs in time  $O(n^3)$ . The overall running time is therefore  $O(\max\{Nn^2, n^3\})$ .

We observe that, in matrix  $\mathbf{K}$ , from its top to the bottom row, if the  $p^{\text{th}}$  row does not contain non-zero elements, then  $p - 1$  is the length of the longest common subsequences in  $\mathcal{R}$ . The following corollaries of Lemma 1 provides algorithms for calculating  $\ell(\mathcal{R})$  and  $\kappa_p(\mathcal{R})$  once  $\mathbf{K}$  is obtained.

**Corollary 1.** Under the assumptions in Lemma 1, the length  $\ell(\mathcal{R})$  of the longest subsequences in  $\mathcal{S}(\mathcal{R})$  can be obtained by

$$\max \{p : \text{there exists } i \in [1, n] \text{ s.t. } K_{pi} > 0\}.$$

**Corollary 2.** Under the assumptions in Lemma 1, the number of all common subsequences of  $\mathcal{R}$  is  $\kappa(\mathcal{R}) = \sum_{1 \leq p \leq \ell(\mathcal{R})} \sum_{i=1}^n K_{pi}$ .

*Proof.* Note that  $\mathcal{S}(\mathcal{R}) = \bigcup_{p=1}^{\ell(\mathcal{R})} \mathcal{S}_p(\mathcal{R})$ , thus  $\kappa(\mathcal{R}) = |\mathcal{S}(\mathcal{R})| = \sum_{p=1}^{\ell(\mathcal{R})} |\mathcal{S}_p(\mathcal{R})| = \sum_{1 \leq p \leq \ell(\mathcal{R})} \sum_{i=1}^n K_{pi}$  by Lemma 1.  $\square$

With  $\kappa_p(\mathcal{R})$ , we are able to give more weight to those longer patterns and penalize the shorter patterns in  $\mathcal{S}(\mathcal{R})$  by using

$$\sum_{p=1}^{\ell(\mathcal{R})} \alpha_p \kappa_p(\mathcal{R}) \quad (7)$$

---

**Algorithm 1:** Pseudo code for calculating  $\mathbf{K}$  (Lemma 1)

---

**Data:** A set of rankings  $\mathcal{R} = \{\mathbf{r}_1, \dots, \mathbf{r}_N\}$   
**Result:**  $\mathbf{K}$

- 1 Pick an arbitrary  $\mathbf{r}_x \leftarrow r_{x_1} \cdots r_{x_n} \in \mathcal{R}$ ;
- 2 Initialize  $\mathbf{M} = (M_{ki})_{N \times n}$ , with  $M_{ki} = \eta_k(r_{x_i})$  ;
- 3 Initialize  $\mathbf{A} = (A_{ij})_{n \times n}$  to be zero matrix;
- 4 **for**  $i \leftarrow 1$  **to**  $n$  **do**
- 5      $A_{ii} \leftarrow \prod_{k=1}^N \delta(M_{ki})$ ;
- 6     **for**  $j \leftarrow 1$  **to**  $i$  **do**
- 7          $A_{ij} \leftarrow \prod_{k=1}^N (M_{ki} - M_{kj}) A_{ii} A_{jj}$  ;
- 8     **end**
- 9 **end**
- 10 Initialize  $\mathbf{K} = (K_{pi})_{n \times n}$  to be zero matrix;
- 11 **for**  $i \leftarrow 1$  **to**  $n$  **do**
- 12      $K_{1i} = A_{ii}$ ;
- 13 **end**
- 14 **for**  $p \leftarrow 2$  **to**  $n$  **do**
- 15     **for**  $i \leftarrow p$  **to**  $n$  **do**
- 16          $K_{pi} \leftarrow \sum_{j=1}^{i-1} K_{p-1,j} A_{ij}$ ;
- 17     **end**
- 18 **end**
- 19 **return**  $\mathbf{K}$

---

where  $0 \leq \alpha_p \leq 1$ ,  $\sum_{p=1}^{\ell(\mathcal{R})} \alpha_p = 1$ .

**Example 1.** Let  $\mathcal{R} = \{\mathbf{r}_1 = bdcea, \mathbf{r}_2 = abcde, \mathbf{r}_3 = bdce\}$ , based on the matrix  $\mathbf{A}$  in Equation (1), we have

$$\mathbf{K}_{5 \times 5} = \begin{matrix} & a & b & c & d & e \\ \begin{matrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{matrix} & \begin{pmatrix} 0 & 1 & 1 & 1 & 1 \\ 0 & 0 & 1 & 1 & 3 \\ 0 & 0 & 0 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{pmatrix} \end{matrix}$$

and  $\ell(\mathcal{R}) = 3$  as the fourth row is the first zero row.

### 3.1 Gap-weighted consensus measure $\hat{\kappa}(\mathcal{R})$

If we examine Figure 1, we can find that pattern  $bc$  and  $be$  are very different in terms of the gaps between them in  $\mathbf{r}_2$ , though both are of length 2. For example, the gap for  $bc$  in  $\mathbf{r}_1$  is 1 as there is only one item  $d$  placed between  $b$  and  $c$ . The following table summarizes the gaps for  $bc$  and  $be$  in the 3 rankings.

	$\mathbf{r}_1$	$\mathbf{r}_2$	$\mathbf{r}_3$
$bc$	1	0	1
$be$	2	2	2

The gaps between two items  $\sigma_i$  and  $\sigma_j$  in rankings suggest the strength of preference for one item  $\sigma_i$  over another item  $\sigma_j$ . The bigger the gap, the less preference of  $\sigma_j$  than that of  $\sigma_i$ . The gap size is an important factor in consensus evaluation, and it makes sense to “penalize” the bigger gaps.

The following lemma makes use of the gap information for evaluating consensus in rankings.

**Lemma 2.** Under the same assumptions in Lemma 1, let  $0 < \lambda \leq 1$ , and  $\hat{\mathbf{K}} = (\hat{K}_{pi})_{n \times n}$ , where

$$\hat{K}_{p,i} = \begin{cases} E_{ii}, & p = 1; \\ \sum_{j=1}^{i-1} \hat{K}_{p-1,j} \lambda^g A_{ij}, & p > 1. \end{cases} \quad (8)$$

and  $g = \sum_{k=1}^N |\eta_k(\sigma_{x_i}) - \eta_k(\sigma_{x_j})|$ .

Then the gap weighted consensus measure  $\hat{\kappa}_p(\mathcal{R})$  with respect to a given length  $p$  is

$$\hat{\kappa}_p(\mathcal{R}) = \sum_{1 \leq i \leq n} \hat{K}_{p,i}$$



and the gap weighted consensus measure  $\hat{\kappa}(\mathcal{R})$  is

$$\hat{\kappa}(\mathcal{R}) = \sum_{1 \leq p \leq \ell(\mathcal{R})} \hat{\kappa}_p(\mathcal{R})$$

Lemma 1 and Corollary 2 are special cases of Lemma 2 when  $\lambda = 1$ . Similar to Equation (7), we can also have  $\sum_{p=1}^{\ell(\mathcal{R})} \alpha_p \hat{\kappa}_p(\mathcal{R})$ .

Table 1: Rankings of top 25 links returned from Google and Bing with the given key words (*BF* for “bond films”, *BM* for “bond movies”, *OM* for “007 movies”, *OF* for “007 films”, *JF* for “james bond films”, and *JM* for “james bond movies”).  $\mathbb{G}$  and  $\mathbb{B}$  stand for Google and Bing respectively. The numbers are the ids for the links, and the mapping of the ids to the links can be find on our github repository.

$\mathbb{G}$	Top 25 links from Google
<i>BF</i>	0,68,9,59,11,5,3,69,79,70,21,4,36,32,76,40,60,51,80,81,42,29,82,83,73
<i>BM</i>	5,0,9,11,59,76,3,36,21,79,90,70,4,60,93,35,50,40,42,92,86,87,73,98,94
<i>OM</i>	0,9,11,5,3,59,70,84,85,21,76,4,32,42,51,62,12,80,67,60,55,86,87,73,29
<i>OF</i>	0,9,3,11,5,2,59,4,32,21,60,35,42,61,62,12,51,17,63,64,55,65,66,40,67
<i>JF</i>	5,0,68,9,3,11,59,69,70,71,4,21,60,32,36,61,65,72,73,58,74,75,76,77,78
<i>JM</i>	0,9,59,5,11,3,88,60,89,69,90,70,32,91,75,92,93,94,61,40,86,87,95,96,97
$\mathbb{B}$	Top 25 links from Bing
<i>BF</i>	0,9,1,11,36,8,4,2,22,5,37,15,38,6,28,34,29,19,39,40,35,24,41,32,42
<i>BM</i>	0,1,9,2,11,5,50,8,4,56,57,22,3,40,51,7,41,15,19,37,36,55,42,58,38
<i>OM</i>	0,7,9,2,10,8,4,6,5,1,11,14,3,40,13,43,19,44,45,46,47,48,49,17,28
<i>OF</i>	0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20,21,22,23,24
<i>JF</i>	9,1,25,4,7,2,0,26,5,11,15,8,27,10,28,29,30,31,6,19,32,22,33,34,35
<i>JM</i>	0,9,3,1,2,11,50,4,8,7,6,38,40,25,10,51,15,19,52,36,53,14,54,55,28

## 4 Experiments

This section presents the results of applying the proposed algorithms to compare the search rankings with Google and Bing. The key words used are “Bond films”, “Bond Movies”, “007 films”, “007 movies”, “James Bond films” and “James Bond movies”, as these key words refer to an identical concept. We generate each ranking with the top 25 unique links from the search engines after each key word is sent. Table 1 shows 12 rankings from the two search engines, in which each search engine has 6 rankings corresponding to

the 6 key words. In the 12 rankings, we obtain altogether 98 distinct links, from which Bing returns 59 distinct links and Google returns 60 distinct links.

The aim of this experiment is to use the proposed measure to evaluate the search results in terms of “relatedness” or “closeness”, with the given conceptually related key words. Table 2 shows the results, from which we can find that the six rankings from Bing search have 8 links in common ( $\kappa_1(\mathcal{R}) = 8$ ) while the Google’s rankings share 7 common links ( $\kappa_1(\mathcal{R}) = 7$ ). The links, which exist in all the six rankings, have been highlighted with boxes in Table 1.

From Table 2, we find that for the given keywords, Google returns more related results than Bing search. For example,  $\ell(\mathcal{R}) = 4$  means that Google’s results have at least a pattern with 4 links, occurring in all its 6 rankings in the same order. When  $p > 1$ , both  $\kappa_p(\mathcal{R})$  and  $\hat{\kappa}_p(\mathcal{R})$  consistently suggest that the 6 rankings from Google has higher consensus than Bing’s rankings.

Table 2: Experimental results with the search rankings (shown in Table 1) from Google  $\mathbb{G}$  and Bing  $\mathbb{B}$  searches.

	$\ell(\mathcal{R})$	$\lambda = 1$				
		$\kappa_1(\mathcal{R})$	$\kappa_2(\mathcal{R})$	$\kappa_3(\mathcal{R})$	$\kappa_4(\mathcal{R})$	$\kappa(\mathcal{R})$
$\mathbb{G}$	4	7	13	10	3	33
$\mathbb{B}$	3	8	11	4	0	23
	$\ell(\mathcal{R})$	$\lambda = 0.95$				
		$\hat{\kappa}_1(\mathcal{R})$	$\hat{\kappa}_2(\mathcal{R})$	$\hat{\kappa}_3(\mathcal{R})$	$\hat{\kappa}_4(\mathcal{R})$	$\hat{\kappa}(\mathcal{R})$
$\mathbb{G}$	4	7	2.857	0.881	0.036	10.77
$\mathbb{B}$	3	8	0.666	0.014	0	8.68

## 5 Conclusion

This paper addresses the consensus measure for a given set of rankings, in order to understand the degree to which the rankings agree and the extent to which the rankings are related. We propose an approach to measuring the consensus in a set of rankings, with respect to the length of common patterns, the number of common patterns for a given length, and the number of all common patterns. We use the search results from Google and Bing with the given key words in the experiment, which shows how the proposed approach can be used to evaluate search quality in terms of closeness and relatedness when conceptually related key words are used.

The future work includes how to extract the common patterns and how to make use the proposed approach for rank aggregation.

## References

- [1] CHAUDHURI, S., DAS, G., HRISTIDIS, V., AND WEIKUM, G. Probabilistic ranking of database query results. In *Proceedings of the Thirtieth International Conference on Very Large Data Bases - Volume 30* (2004), VLDB '04, VLDB Endowment, pp. 888–899.
- [2] ELZINGA, C. H., WANG, H., LIN, Z., AND KUMAR, Y. Concordance and consensus. *Information Sciences* 181, 12 (2011), 2529 – 2549.
- [3] KENDALL, M. G. A new measure of rank correlation. *Biometrika* 30, 1/2 (1938), 81–93.
- [4] MANNING, C. D., RAGHAVAN, P., AND SCHÜTZE, H. *Introduction to Information Retrieval*. Cambridge University Press, New York, NY, USA, 2008.
- [5] QUESADA, F. J., PALOMARES, I., AND MARTÍNEZ, L. Managing experts behavior in large-scale consensus reaching processes with uninorm aggregation operators. *Appl. Soft Comput.* 35, C (Oct. 2015), 873–887.
- [6] SPEARMAN, C. The proof and measurement of association between two things. *The American Journal of Psychology* 15, 1 (1904), 72–101.